



Techniques for Making Molehills out of Unstructured Data Mountains

Visual analytics, a new technology that graphically illustrates datasets, helps users quickly identify responsive documents for electronic discovery review.

Kevin Carr

The astounding volume of data produced, shared, and stored by organizations today is accelerating at a greater pace than ever before. Managing this information in the past posed some challenges, but what was once considered “a lot of data” is nothing compared to what is now measured in terabytes (1,000 gigabytes) or even petabytes (1 million gigabytes).

Prior to when most data was created and stored electronically, business professionals would create a document, use it for its intended purpose, and then periodically make decisions about whether or not to file the information. Organizations

At the Core

This article

- ▶ Describes visual analytics, a technology developed to help users cull large collections of data
- ▶ Illustrates how visual analytics enables users to more quickly evaluate the contents of a dataset and identify the relevant data
- ▶ Tells how visual analytics can help organizations save time and money, protect intellectual property, and limit liability

archived only that which they deemed truly important because they had neither the time nor the money to engage in elaborate document storage systems.

With the adoption of and increased reliance on computers, the decision to retain information no longer revolves around manually filing a document; it focuses on actively deleting it. But with the availability of petabytes of computer storage, workers may not feel the need to delete or destroy files. Predictably, organizations have amassed huge volumes of archived materials, saved on hard drives or back-up tape media.

Over time, offsite storage of archived documents has become the norm.

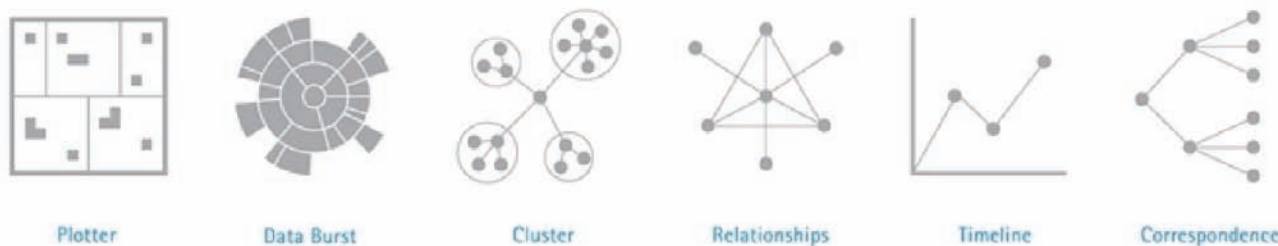


Figure 1: Visual Analytics' Graphical Representations of Data

Data mapping with visual analytics allows the user to see graphical representations of data.

However, with the materials now stored remotely, an “out-of-sight, out-of-mind” approach to dealing with the data also has become common. As a result, organizations often find themselves overwhelmed when required to sort through the data pool to produce responsive documents during litigation or regulatory compliance activities in preparation for electronic discovery review. Collecting all of this data takes a great deal of time, requiring a number of steps, often starting with tape restoration. A series of processing activities follows, including de-duplication, keyword searches, and data filtering, each of which takes times and may add thousands of dollars in associated expenses.

Adding to the frustration over these mountains of data that must be managed is the reality that only a small portion of each document collection is even responsive to the case at hand. So, after dedicating significant time and expense to collecting and processing vast amounts of data, much of the effort is inevitably for naught.

The good news is that today’s technology offers some help in dealing with large sets of unstructured data, with some tools taking a very logical approach to leverage the strengths of both man and machine. One such advancement is the development of a visual method for analyzing and managing data collections.

Pictures Are Worth a Thousand Words

Research has shown that people generally tend to be visual in nature, and – given the choice – they prefer to view graphical

or illustrative representations of material as opposed to text. Not only is it their preference to receive information in this format, but people typically tend to process visually presented information faster and are more inclined to retain it.

Visual techniques for learning and processing a wide range of information are known to be quite effective with a vast majority of the general population. This is why teachers, for example, use a variety of visual aids in their classrooms and why lawyers use video and other graphical illustrations in their trial presentations.

But can images or any other kind of visual demonstration be effectively used to manage and analyze vast collections of data? The latest technology garnering significant attention in the legal industry leverages the visual nature of the human mind to do just that in a way that is revolutionizing how attorneys develop their discovery and case strategies.

Visual analytics is a new approach to reviewing large collections of data. In fact, this method of analyzing the contents of a dataset is among the best available for collections of significant size because its effectiveness is not compromised regardless of how many documents might be included. Applying visual analytics to a mountain of data is one of the fastest, easiest, and most comprehensive ways of culling that data to a more manageable, molehill-sized set.

Mapping the data in this way also promotes a quicker understanding of the collection’s contents from a broad perspective, which helps shape the strategy that will be employed for the duration of the discovery

life cycle. Without having to invest significant time to achieve this intelligence, attorneys can know what they’re working with earlier in the case than ever before.

How Visual Analytics Works

When evaluating a dataset for litigation or a regulatory compliance matter, for example, the traditional approach is to view a linear listing of documents that must be viewed one-by-one to determine relevance and responsiveness to the case. The need to look at documents during a review process is unlikely to change in the foreseeable future. However, there are newer ways to evaluate the contents of the dataset earlier in the process that allow reviewers to focus on the documents that are most important to the case and to eliminate those that have no bearing. This not only helps with formulating a successful case strategy, but it can save a tremendous amount of time and money with e-discovery processing and review.

While today’s technology provides for the quick detection of duplicate and near-duplicate documents to winnow the entire collection to some extent, other search technologies must then be employed to find documents that share something in common – such as keywords or concepts. However, even once those are identified, what is delivered is generally a list of documents.

Applying visual analytics to the data provides graphical illustrations of search results in addition to this linear listing. These graphical illustrations portray the data in a variety of models, including basic

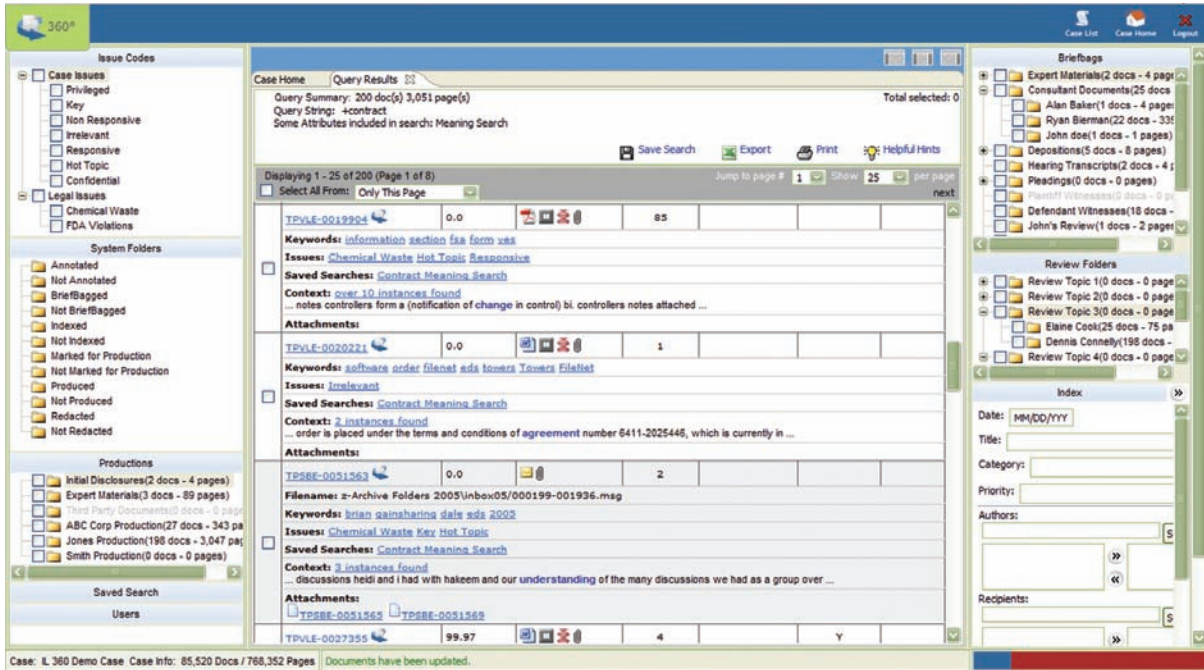


Figure 2: Search Results for Linear Review

In addition to offering concept and visual mapping tools, visual analytics can also provide a “flat list” search results environment for instances in which a linear review of a set of documents is needed. More advanced technologies can be used to cull down to a set much more manageable for reviewers to do a quick document-by-document review.

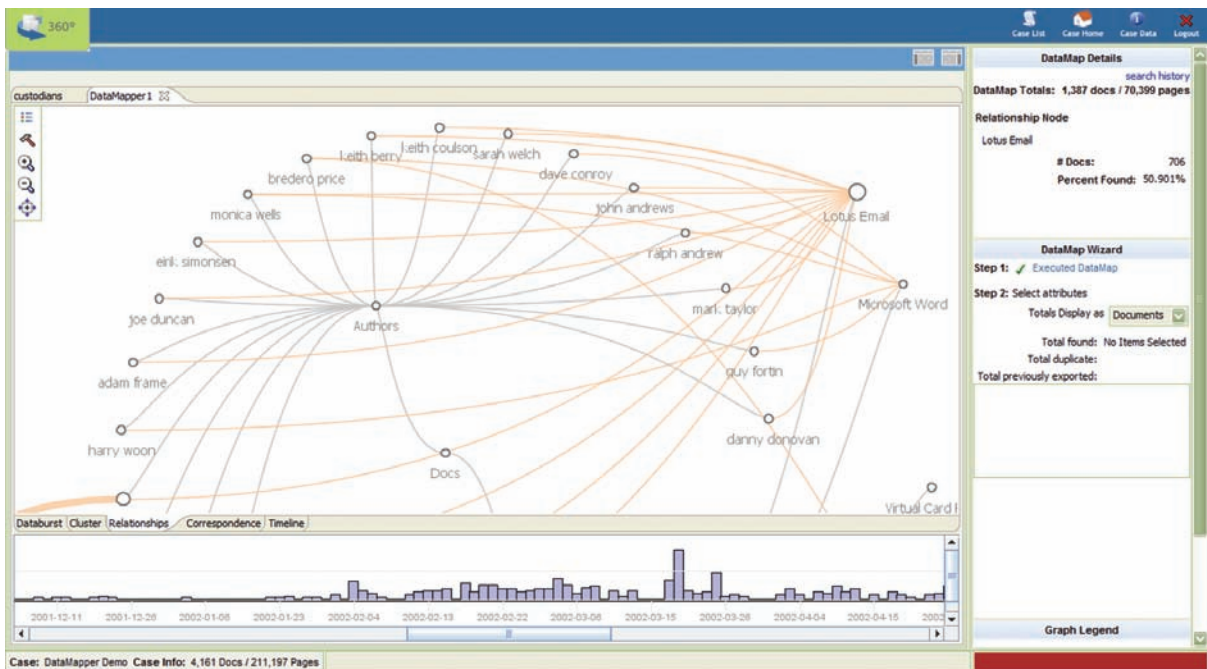


Figure 3: Relationships Among Documents

As part of visual analytics graphing, users can map out all relationships among documents. This enables users to see all the “common threads” of similarity among documents using any set of criteria legal teams feel is relevant.

of data that have no relevance and can be eliminated from the collection entirely.

This negates the need to spend time or money having these documents initially processed – and reviewed later on. Pre-culling data collections is a topic all on its own; it is one that has legitimately gained tremendous attention in recent months. Any time a dataset can be culled to reduce the quantity of irrelevant documents and keep the focus on those that matter, everybody wins.

Gaining Control

Some organizations are bringing certain activities that were once outsourced back in-house in an effort to maintain greater control over their data and the overall process. Visual analytics technology helps with this initiative as well.

Nobody knows an organization's information as well as that organiza-

tion's employees – not its outside counsel and certainly not other third-party service providers. Certain applications offering visual analytics technology are designed with the data owners in mind; they are intuitive and most efficient when used by the organization itself.

By using visual tools to cull a data collection prior to sharing with outside counsel or any vendors, the organization minimizes its exposure to potential risks posed by the unnecessary production of intellectual property or the inadvertent disclosure of privileged data.

Whether used early in a compliance or

litigation matter to pre-cull a large, unstructured collection of data in an effort to achieve a more manageable dataset, or to perform a more in-depth review of the entire collection of documents, visual analytics technology offers a unique, pictorial view of the information at hand. Many of these tools haven't been available until just recently, but the methodology behind them has been used in various capacities for a very long time. The technology simply leverages the visual nature of the human mind to better manage what might otherwise be an overwhelming project. ■

Kevin Carr is president of InterLegis Inc., Dallas, Texas, and has a wide range of Internet-based technology expertise. He is the architect of his company's visual analytics technology, which works with corporate legal departments, law firms, and legal service providers to reduce the risk, complexity, and cost of litigation, regulatory requirements, and internal investigations through early case assessment, electronic data culling and processing, hosted review, and advanced analytics. He may be reached at kcarr@interlegis.com.